University of Heidelberg

Department of Economics

Discussion Paper Series | No. 423

Rage Against the Machines

Peter Dürsch, Albert Kolb,
Jörg Oechssler and
Burkhard C. Schipper

Oktober 2005

Rage Against the Machines:
How Subjects Learn to Play Against Computers[*]

Peter Dürsch[§]   Albert Kolb[#]   Jörg Oechssler[§]

Burkhard C. Schipper[♮]

October 12, 2005

Abstract

We use an experiment to explore how subjects learn to play against computers which are programmed to follow one of a number of standard learning algorithms. The learning theories are (unbeknown to subjects) a best response process, fictitious play, imitation, reinforcement learning, and a trial & error process. We test whether subjects try to influence those algorithms to their advantage in a forward-looking way (strategic teaching). We find that strategic teaching occurs frequently and that all learning algorithms are subject to exploitation with the notable exception of imitation. The experiment was conducted, both, on the internet and in the usual laboratory setting. We find some systematic differences, which however can be traced to the different incentives structures rather than the experimental environment.

JEL codes: C72; C91; C92; D43; L13.
Keywords: learning; fictitious play; imitation; reinforcement; trial & error; strategic teaching; Cournot duopoly; experiments; internet.

1 Introduction

In recent years, theories of learning in games have been extensively studied in experiments. The focus of those experiments was mainly on which learning theory describes the average behavior of subjects best. It turned out that some very simply adaptive procedures like reinforcement learning, best response dynamics, or imitation were fairly successful in describing average learning behavior of subjects.

The focus of the current experiment is different. First, we are interested in the strategic aspect of learning in games. Given my opponent plays according to a some learning theory, how should I respond? In the spirit of Nash equilibrium, one can ask whether a learning theory is a best response to itself. Otherwise, it will probably not be sustainable. A second closely related aspect is the evolutionary perspective. Given a population in which everyone uses a given learning theory, could a player endowed with some other learning theory enter the population and be successful?

These two aspects of learning in games have received some attention in the theoretical literature. While Matros (2004) and Schipper (2004) address the evolutionary selection of learning theories, Ellison (1997) and Fudenberg and Levine (1998) deal with the strategic aspect of learning. For example Fudenberg and Levine (1998, p. 261) write "A player may attempt to manipulate his opponent's learning process and try to "teach" him how to play the game. This issue has been studied extensively in models of "reputation effects," which typically assume Nash equilibrium but not in the context of learning theory." Following Camerer and Ho (2001) and Camerer, Ho, and Chong (2002) we shall call this aspect of learning "strategic teaching". We believe that this hitherto largely neglected aspect of learning is of immense importance and deserves further study. As we shall see in this experiment, theories just based on adaptive processes will not do justice to the behavior of subjects.

To address those questions in an experiment, it would be desirable to identify subjects who consistently play according to some learning theory and check how other players react to this. Prior experiments (see e.g. Huck,

Normann, and Oechssler, 1999) suggest, however, that few subject follow a given learning theory with high consistency. Therefore, we decided to let subjects play against computers programmed with particular learning theories. Subjects know that they play against computers.

We consider five learning theories in a Cournot duopoly: best-response (br), fictitious play (fic), imitate-the-best (imi), reinforcement learning (re), and trial & error (t&r). Some noise is added in order to make the task less than obvious. Noise is also a requirement for some of the theoretical predictions to work as it prevents a learning process from getting stuck at states which are not stochastically stable.[1] A Cournot duopoly is chosen because of its familiarity in theory and experiments. The selection of learning theories is based on three criteria: (1) prominence in the literature, (2) convenient applicability to the Cournot duopoly, and (3) sufficient variety of theoretical predictions.

The experiment was conducted on the internet as well as in a traditional laboratory environment. Internet experiments are still relatively novel (see e.g. Drehmann, Oechssler, and Roider, 2005, for first experiences). Arguably, the setting (working at home or in the office at your own PC) is more representative of real world decisions than in the usual laboratory experiments. On the other hand, experimenters lose control to some extent, and many methodological questions are still unsettled. That is why we also run a control experiment in the usual *lab* setting.

Our design allows us to address questions such as: How well do subjects do against computers programmed to various learning theories? Do subjects try to strategically teach computers, and if so how? Can the same learning theories, which were used to program the computers, also describe the subjects' behavior? We find that strategic teaching occurs frequently and that all learning algorithms are subject to exploitation with the notable exception of imitation. This primarily shows up in the fact that human subjects achieve substantially higher profits than those learning algorithms. As expected from the theoretical analysis (see e.g. Schipper, 2004), the exception

---

[1]See e.g. Vega-Redondo (1997) for imitate-the-best and Huck, Normann, and Oechsler (2004a) for trial & error.

is the imitation algorithm, which cannot be beaten by more than a small margin and which performs on average better than its human opponents. On the other hand, our subjects learned quickly how to exploit the computers programmed to best response and fictitious play, usually by behaving as Stackelberg leader, although some subjects managed to find more innovative and even more profitable ways. The computer opponent that allowed the highest profits for its human counterparts was the reinforcement learning computer. However, due to stochastic nature of reinforcement learning, a lot of luck was needed and variances were high.

We also compare our data to a similar experiment in which, as usual, human subjects played against human subjects. This comparison yields some interesting differences. Human subjects are much less aggressive against other human subjects than against computer opponents. When computers are more accommodating (i.e. when they are programmed to follow best response or fictitious play) this increase in aggressiveness yields higher profits. The opposite happens, when the computer is programmed to play imitation. In that case both competitors have very low profits or even suffer losses.[2]

There is already a small literature on experiments where subjects play against computers. Most of this literature is concerned either with mixed-strategy equilibrium in zero-sum games or with controlling for social preferences or fairness considerations. Lieberman (1961), Messick (1967), and Fox (1972) found that subjects are not very good in playing their minimax strategy against a computer opponent which plays its minimax strategy in zero-sum games. Shachat and Swarthout (2002) let subjects play against both, human subjects and computers, which are programmed to follow reinforcement learning or experienced weighted attraction in repeated 2x2 games with a unique Nash equilibrium in mixed strategies. They found that human play does not significantly vary depending on whether the opponent is a human or a programmed learning algorithm. In contrast, the learning algorithms respond systematically to non-Nash behavior of human subjects.

---

[2] This is further evidence that imitation yields very competitive outcomes. See Vega–Redondo (1997) for the theoretical argument and Huck, Normann, and Oechssler (1999) and Offerman, Potters, and Sonnemans (2002) for experimental evidence.

Nevertheless, these adjustments are too small to result in significant payoff gains. Coricelli (2001), on the other hand, found that human subjects do manage to exploit computer opponents that play a biased version of fictitious play in repeated 2x2 zero-sum games.

Walker, Smith, and Cox (1987) used computerized Nash equilibrium bidders in first price sealed bid actions. They found no significant difference in subjects' bidding whether they play against computers or human subjects (subjects knew when they were playing against computers). In contrast, Fehr and Tyran (2001) found a difference in subjects' behavior in a money illusion experiment depending on whether subjects played against computers or against real subjects.[3]

Roth and Schoumaker (1983) used computer opponents to control for expectations of subjects in bargaining games. Kirchkamp and Nagel (2005) used computer players to plant a "cooperative seed" in a local interaction model where subjects play a prisoner's dilemma.

McCabe et al. (2001) showed using brain imagining techniques that the prefrontal cortex is relatively more active when subjects play against humans than against programmed computers in a trust game. This was less pronounced for subjects that choose mostly non-cooperatively. It is speculated that the prefrontal cortex is connected to trading off immediate gratification and mutual gains. Finally, Houser and Kurzban (2002) used programmed computers to control for social motives in a public goods experiment.

The remainder of the paper is organized as follows. Section 2 describes the Cournot game that is the basis for all treatments. In Section 3 we introduce the computer types and the associated learning theories. The experimental design is explained in Section 4, followed by the results in Section 5. Subsection 5.6 discusses the differences between the internet and the laboratory setting. Section 6 concludes. The instructions for the experiment and screenshots are shown in the Appendix.

---

[3]However, Fehr and Tyran told their subjects which rule the computer used. Thus, in contrast to the treatment with real subjects, there was no strategic uncertainty.

2   The Cournot game

We consider a standard symmetric Cournot duopoly with linear inverse demand function $\max\{109 - Q, 0\}$ and constant marginal cost of 1. Each player's quantity q_i, $i = 1, 2$ is an element of the discrete set of actions $\{0, 1, ..., 109, 110\}$. Player i's profit function is given by

$$\pi(q_i, q_{-i}) := (\max\{109 - q_i - q_{-i}, 0\} - 1) \, q_i. \tag{1}$$

Given this payoff function it is straightforward to compute the Nash equilibrium and several other prominent outcomes like the symmetric competitive outcome, the symmetric collusive outcome, the Stackelberg leader and follower outcomes, and the monopoly solution. See Table 1 for the corresponding output and profit values.

Table 1: Prominent outcomes

	q_i	q_{-i}	π_i	π_{-i}
Cournot Nash equilibrium	36	36	1296	1296
symmetric competitive outcome	54	54	0	0
symmetric collusive outcome	27	27	1458	1458
Stackelberg leader outcome	54	27	1458	729
Stackelberg follower outcome	27	54	729	1458
monopoly solution	54	0	2916	0

Subjects play the Cournot duopoly repeatedly for 40 rounds. Thus, we index the quantity q_i^t by the period $t = 1, ..., 40$.

3   Computer types

Computers were programmed to play according to one of the following decision rules: Best-response (br), fictitious play (fic), imitate the best (imi), reinforcement learning (re) or trial & error (t&r). All decision rules except

reinforcement learning are deterministic, which would make it too easy for subjects to guess the algorithm (as we experienced in a pilot study to this project). Therefore, we introduced some amount of noise for the deterministic processes (see below for details). The action space for all computer types was $\{0, 1, ..., 109\}$.

All computer types require an exogenously set choice for the first round as they can only condition on past behavior of subjects. To be able to test whether starting values matter, we chose different starting values. However, to have enough comparable data, we restricted the starting values to 35, 40, and 45. Starting quantities were switched automatically every 50 subjects in order to collect approximately the same number of observations for each starting quantity but subjects were unaware of this rule.

3.1   Best-response (br)

Cournot (1838) himself suggested a myopic adjustment process based on the individual best-response

$$q_i^t = \arg\max_{q_i} \pi(q_i, q_{-i}^{t-1}) = \max\left\{\frac{108 - q_{-i}^{t-1}}{2}, 0\right\}, \tag{2}$$

for $t = 2, ...$ Note that there is a unique best response for each opponent's quantity choice. Moreover, the parameters are such that if both players use the best-response process, the process converges to the Nash equilibrium in a finite number of steps (see e.g. Monderer and Shapley, 1996). This holds for both, the simultaneous version of the process (when both players adjust simultaneously) and the sequential version (when only on of the players adjusts quantities every period).

This deterministic process is supplemented by noise in the following way. If the best response process yields some quantity q_i^t, the computer actually plays a quantity chosen from a Normal distribution with mean q_i^t and standard deviation 2, rounded to the next integer in $\{0, 1, ..., 109\}$.[4]

---

[4]Due to a programming error in the rounding procedure, the noise was actually slightly biased downwards (by 0.5), which makes the computer player slightly less aggressive. This does not have any lasting effects for computer types br and fic but has an effect on imi.

6

This implementation of noise is also used for computer types fictitious play and imitation.

3.2 Fictitious play (fic)

A second decision rule that is studied extensively in the literature is fictitious play (see Brown, 1951, Robinson, 1951, and Fudenberg and Levine, 1998, chapter 2). A player who uses fictitious play chooses in each round a myopic best response against the historical frequency of his opponent's actions (amended by an initial weight for each action). If we let those initial weight be the same for each action and each player, $w_i^0(q_{-i}) = w^0$, we obtain the following recursive formulation for the weight player i attaches to his opponent's action q_{-i}, where 1 is added each time the opponent chooses q_{-i}.

$$w_i^t(q_{-i}) = w_i^{t-1}(q_{-i}) + \begin{cases} 1 & \text{if} \quad q_{-i}^{t-1} = q_{-i} \\ 0 & \text{if} \quad q_{-i}^{t-1} \neq q_{-i} \end{cases}$$

for $t = 2, ...$ Player i assigns probability

$$p_i^t(q_{-i}) = \frac{w_i^t(q_{-i})}{\sum_{q'_{-i}} w_i^t(q'_{-i})}$$

to player $-i$ using q_{-i} in period t. Consequently, player i chooses a quantity that maximizes his expected payoff given the probability assessment over the opponent's quantities, i.e.,

$$q_i^t \in \arg\max_{q_i} \sum_{q_{-i}} p_i^t(q_{-i})\pi(q_i, q_{-i}). \tag{3}$$

We simulated the fictitious play processes against itself and some other decision rules for many different initial weighty w^0 and ended up choosing $w^0 = 1/25$. Except for much smaller or much larger initial weights, results of the simulations did not change much. Very high initial weights lead to rather slow adaptation whereas very small ones resulted in erratic movements. Since our Cournot duopoly is a potential game, fictitious play must converge to the unique Cournot Nash equilibrium (see Monderer and Shapley, 1996).

3.3 Imitate the best (imi)

Imitation has received much attention recently in both theory and experiments (see e.g. Vega-Redondo, 1997, Apesteguia et al. 2004, Schipper, 2004). The rule "imitate the best" simply requires to choose the best action that was observed in the previous period. If player i follows this decision rule in $t = 2, ...,$ he chooses

$$q_i^t = \begin{cases} q_i^{t-1} & \text{if} \quad \pi(q_i^{t-1}, q_{-i}^{t-1}) \geq \pi(q_{-i}^{t-1}, q_i^{t-1}) \\ q_{-i}^{t-1} & \text{otherwise.} \end{cases} \tag{4}$$

Vega-Redondo (1997) shows for symmetric Cournot oligopoly that if players follow this decision rule up to a small amount of noise, then the long run distribution over quantities assigns probability 1 to the competitive outcome. The reason is that if a player deviates to the competitive outcome, then he may reduce his profits but reduces the profits of the other player even more. Consequently he will get imitated in subsequent periods.

Schipper (2004) shows that if there are both imitators and best-response players in the game, then any state where imitators are weakly better off than best-response players and where best-response players play a best-response is absorbing. Moreover, if mistakes are added, then in the long run imitators are strictly better off than best-response players. The intuition is that if imitators play a sufficiently large quantity, best-responders become Stackelberg followers. Moreover, imitators do not change because they are better off than best-responders.

Alos-Ferrer (2004) shows that if imitators take a finite number past periods into account when deciding on this period's quantity, then the support of the long run distribution contains all symmetric combinations of quantities between the Cournot Nash equilibrium and the competitive outcome. The intuition is that imitators increasing their relative payoff may remember that they had a higher payoff with a different quantity several periods ago. Consequently they will return improving their absolute profits even though they reduce their relative profits.

3.4   Reinforcement learning (re)

Ideas of reinforcement learning have been explored for many years in psychology (e.g. Thorndike, 1898). Roth and Erev (1995) introduced a version of it to games based on the *law of effect*, i.e., choices with good outcomes in the past are likely to be repeated in the future, and the *power law of practice*, i.e., the impact of outcomes decreases over time.

In the standard model of Roth and Erev (1995), an action is chosen with probability that is proportional to the *propensity* for this action. Propensities, in turn, are simply the accumulated payoffs from taking this action earlier in the process.

In games with a large action space such as a Cournot duopoly, it seems unreasonable to reinforce only that single action that was chosen in a given round. Rather, actions in the neighborhood should also be reinforced although to a lesser extent depending on their distance to the original choice. We follow the standard model of reinforcement learning by Roth and Erev (1995) but complement it with updating of neighborhoods a là Sarin and Vahid (2004).

The player starts with an initial propensity for each quantity, $w_i^0(q)$ for all $q \in A$ and $i = 1, 2$. Let q^{t-1} be the quantity chosen in period $t - 1$, $t = 2, ...$ Then propensities are updated by

$$w_i^t(q) = w_i^{t-1}(q) + \beta(q, q^{t-1})\pi_i(q^{t-1}, \cdot),$$

where β is the linear *Bartlett function*

$$\beta(q, q^{t-1}) := \max\left\{0, \frac{6 - |q - q^{t-1}|}{6}\right\}.$$

That is, all actions within 5 grid points of the chosen action are also reinforced.

The probability of playing quantity q in period t is computed by normalizing the propensities

$$p_i^t(q) = \frac{w_i^t(q)}{\sum_{q'} w_i^t(q')}.$$

9

Theoretical results on the convergence properties of reinforcement learning are scarce.[5] Thus most of the analysis is based on simulations. We ran several simulations of reinforcement learning against itself as well as other decision rules while varying the initial propensities $w_i^0(q)$. Results did not change much when using different initial propensities. We chose $w_i^0(q) = 78$, which minimized the mean squared deviation to the Nash equilibrium. Since reinforcement learning already is a stochastic process, we did not add additional noise to the process.

3.5   trial & error (t&e)

Huck, Normann and Oechssler (2004a) introduce a very simple trial & error learning process. Players begin by adjusting their initial quantity either up- or downwards with an exogenously fixed step size. If this change increases profits, the direction is continued. If it does not, the direction of adjustment is reversed. We chose a step size of 4. Formally, player adjust their quantities as follows:

$$q_i^t := \max\{0, \min\{q_i^{t-1} + 4s_i^{t-1}, 109\}\},$$

for $t = 2, ...,$ where

$$s_i^t := \begin{cases} sign(q_i^t - q_i^{t-1}) \times sign(\pi_i^t - \pi_i^{t-1}) & \text{if} \quad (q_i^t - q_i^{t-1})(\pi_i^t - \pi_i^{t-1}) \neq 0 \\ +1, -1 \text{ each with positive probability otherwise.} \end{cases}$$

On the boundaries of the output grid, we chose a "soft reflecting boundary". In particular, when a player repeated 109 or 0 twice in subsequent periods, the next quantity chosen was $109 - 4$ or $0 + 4$, respectively.

Huck, Normann and Oechssler (2004a) show that in Cournot duopoly if players are allowed to choose the wrong direction with small but positive probability, then trial & error learning converges in the long run to a set of outcomes around the collusive outcome. To follow the theoretical setting, the noise for this process was modelled such that the computer chose

---

[5]Laslier, Topol and Walliser (2001) show that reinforcement learning converges with positive probability to any strict pure Nash equilibrium in finite two-player strategic games. Similar results were obtained by Ianni (2002). However, they do not consider reinforcement of neighborhoods as in our case.

the opposite direction from that prescribed by the theory with independent probability of 0.2 in each round.

4 Experimental design

More than 600 subjects participated in our experiment. The bulk of the experiment was conducted as an internet experiment (setting *net*). Additionally there was a control experiment conducted as a regular laboratory experiment with the usual monetary incentives (setting *lab*). In *net*, subjects played on the internet, in a location of their own choice (home, office etc.), and at their own pace. Recruitment was done by email, newsgroups (like sci.econ, sci.math, sci.psych etc.), and a University of Bonn student magazine. Each recruitment announcement contained a different hyperlink such that we were able to differentiate between subject pools depending on where they were recruited. Each subject chose her/his nickname. On the internet, incentives were provided exclusively by publicly displaying a highscore after the experiment (like in computer games).

In setting *net*, subjects could repeat the experiment as often as they desired, either immediately or at some later time. Subjects were encouraged to repeat under the same user name as before.[6]

In setting *lab*, subjects played in the Bonn Laboratory for Experimental Economics. Subjects were required to repeat the experiment once with the same computer type as opponent, i.e., they played two times 40 rounds as outlined above. Since there were fewer observations in the lab, we used only a starting value of 40 for the computer types. Incentives were provided by paying subjects immediately at the end of the experiment the sum of profits over all rounds according to an exchange rate of 9000 Points to 1 Euro. On average, subjects earned 10.17 Euros for about half an hour in the lab. The instructions for both settings were the same up to the incentive structure (highscore in *net*, cash payment in *lab*).

---

[6]The incentives for doing so were the highscore and the possibility to pick the same computer opponent as before (subjects logging in under a different name were allocated to a randomly chosen computer). The latter possibility was only revealed once subjects logged in under the same name.

The sequence of events was as follows. After logging in (after entering the lab, respectively), subjects were randomly matched to a computer type. The computer type was displayed to subjects via a label (Greek letters) though subjects were not told how computer types were associated with labels. In the instructions (see Appendix A) subjects were told the following: "The other firm is always played by a computer program. The computer uses a fixed algorithm to calculate its output which may depend on a number of things but it cannot observe your output from the current round before making its decision."

A page with instructions was displayed to subjects. At any time during the experiment, subjects were able read the instructions and an example for calculating profits by opening a separate window on their computer. After reading the instructions, subjects could input their quantity for the first round. The computer displayed a new window with the results for the current round including the number of the round, the subject's quantity, the subject's profit, the computer's quantity as well as the computer's profit (see Appendix B for screenshots). A subject had to acknowledge this information before moving on to the following round. Upon acknowledgment, a new page appeared with an input field for the new quantity. This page also showed a table with the history of previous round(s)'s quantities and profits for both players.

After round 40, subjects were asked to fill in a brief questionnaire (see Appendix) with information on gender, occupation, country of origin, formal training in game theory or economic theory, previous participation in online experiments, and the free format question "Please explain in a few words how you made your decisions". It was possible to skip this questionnaire. The highscore was displayed on the following page. This table contained a ranking among all previous subjects, separately for subjects who were matched against the same computer type and for all subjects. It also contained the computer's highscore.

In both the *net* and the *lab* setting, subjects were able to see the entire history from the previous rounds. In an additional internet setting called "no history" (*noh*) we restricted this information to that from the previous

period. This should be informative as some learning theories condition only on the previous round whereas others use the entire history. Table 2 provides a summary of the three experimental settings. Given the three settings and the five learning theories (and neglecting the 3 different starting quantities for the computer), we have 15 treatments.

Table 2: Summary of experimental settings

setting	recruitment	repetition	incentives	history	computer's initial quantity
net	newsgroups	possible	highscore	full	35, 40, 45
lab	laboratory	twice	profit	full	40
noh	newsgroups	possible	highscore	previous round	35, 40, 45

The experiments were conducted in November 2003 in the Bonn Laboratory of Experimental Economics and from December 2003 until March 2004 on the internet. Table 3 lists the number of first time players and the number of repeaters for each setting. Recall that subjects in the internet setting were allowed to repeat as often as they liked.[7]

Table 3: Number of subjects

	first–timer	repeater
net	550	500
noh	81	30
lab	50	50
total	681	580

The technical implementation of the experiment was based on the following criteria: (1) easy access, (2) minimal technical requirements, (3) high system stability, and (4) high system security. In order to participate in our experiment, a standard web browser and a low-speed internet connection were sufficient. That is, no plug-ins like Flash or ActiveX Object or technologies such as cookies or JavaScript were required. We did not want to exclude (and implicitly select) subjects by technical means. To separate among different subject pools, we used different virtual directories. Each

---

[7]The record was a subject who played 31 times.

subject pool (e.g. different newsgroups) was informed of a different link, and subjects were unaware of other links.

Our servers were based on Windows Server 2003. We used IIS 6.0 with ASP-technology as the web-based solution as well as Microsoft SQL 2000 SP3 as database. This technology allows for easy back-up, remote-access, failure diagnostics, and a standardized SQL-to-SPSS interface.

5   Results

To give a first impression of the data, we present in Table 4 mean quantities of subjects and computers, respectively, averaged over all rounds and subjects. The first thing to notice is that subjects on average have much higher quantities than computers (47.95 vs. 34.39). This holds for all treatments except for the imitation treatments. Recall that the Cournot–Nash quantity is 36 (see Table 1). Thus, subjects chose on average quantities that exceed by far the Cournot quantity and in some cases come close to the Stackelberg leader output of 54.

A further observation is that quantities in the lab seem to be generally lower than on the net. We will comment on this difference in Section 5.6. Average quantities for the no history setting (*noh*) are also somewhat lower than for *net*. At a first glance, this is surprising because some learning theories predict, if anything, the opposite (e.g. imitation with a 1-period memory yields more competitive outcomes than imitation with longer memories, see Alos–Ferrer, 2004). However, the data corresponds nicely to our evidence on strategic teaching (see Section 5.2 below). Strategic teaching is probably easier to do if one has available a longer track record of the computer's quantities. And since strategic teaching, in most cases, leads to more aggressive play in a Cournot game, this would explain the finding.

5.1   How do subjects do against computers?

In the end, what matters are subjects' profits. How do they differ with respect to the different computer types? Figure 1 report the range of subjects' average profits per round and mean profit per round of first time players and

Table 4: Mean quantities

treatment	subjects' mean quantities	computers' mean quantities
br_net	51.99 (10.80)	27.79 (5.25)
br_lab	48.67 (9.25)	29.34 (4.58)
br_noh	49.18 (12.08)	29.23 (5.80)
t&e_net	48.96 (9.92)	32.05 (6.91)
t&e_lab	38.49 (4.20)	35.02 (3.99)
t&e_noh	45.90 (7.97)	31.67 (5.68)
fic_net	46.11 (10.15)	31.94 (3.53)
fic_lab	41.27 (5.47)	33.82 (2.49)
fic_noh	43.62 (6.94)	32.71 (2.63)
imi_net	46.40 (11.39)	48.38 (6.11)
imi_lab	40.29 (7.22)	45.37 (6.67)
imi_noh	45.92 (7.23)	49.57 (6.58)
re_net	47.45 (11.67)	35.71 (10.08)
re_lab	42.80 (7.50)	37.64 (10.34)
re_noh	45.71 (17.74)	43.55 (15.07)
Total	47.95 (10.96)	34.39 (9.55)

Note: Average quantities over all 40 rounds and all subjects in a given treatment. The Cournot-Nash equilibrium quantity is 36. Standard deviations in parentheses.

repeaters, respectively. The figures report those measures separately for each of our treatments, i.e. for each combination of computer type (br, t&e, fic, imi, and re) and setting (*net, lab, noh*). The dotted line indicates the profit per round in the Cournot Nash equilibrium.

First time players who are matched with a computer types br, t&e, or fic achieve on average slightly less than the Nash equilibrium profit. The ranges in profits are larger in the internet treatments than in the lab but roughly comparable across the three computer types. Drastically different, however, are profits of subjects who were matched against the computer types imi and re. On average profits against imi were less than half the profits against the first three computer types. Even the very best subjects do not reach the Nash equilibrium profit, despite the bias in the noise of this computer type (see Footnote 4). Profits against computer type re are also substantially lower than against br, t&e, or fic but they are higher than

against imi.[8] The range of profits is highest against this type of computer. Some subjects achieve very high profits that exceed the Stackelberg leader or collusive profit (of 1458).

Average profits of repeaters are generally higher than those of first time players. The improvements, however, seem to be more pronounced for the internet treatments where subjects could repeat several times and had the choice of computer opponent. While subjects improve somewhat against computer type imi, average payoffs are still by far the lowest of all computer types. Against br and fic, subjects on average do better than the Nash equilibrium profit. The very best subjects played against t&e and re on the *net*.

It is also quite instructive to consider average profits over time. Figure 2 shows profits (averaged over settings *net, lab, noh* and all subjects) of subjects and computers for all 40 periods. Subjects playing against type br almost immediately gain a substantive edge over the computer and keep their profits more or less constant somewhere between the Stackelberg leader profit and the Nash equilibrium profit. The final result against type fic is similar but convergence is much more gradual. The fictitious play computer is also the most successful among the computer types as it stabilizes at a profit of above 1000. The learning curves against types t&e and re look similar, although against the latter subjects do not even manage to achieve the Nash equilibrium profit on average.[9]

A totally different picture yields computer type imi. In contrast to all others, payoffs against imi decrease over time, both for subjects and for computers. Furthermore, it is the only computer type where subjects' payoffs are lower than those of computers. We say more on this below.

If we consider the overall top subjects a slightly different picture emerges (see Table 5). Among the top 100 subjects there are 52 subjects who played

---

[8]For first-time players, profits against re are lower than against br, fic, and t&e according to two–sided MWU tests at $p < 0.01$. For repeaters only the first difference remains significant at $p = 0.02$. For both, first-timers and repeaters, profits against re are higher than against imi at $p < 0.001$.

[9]The dip of the computer player in round 2 is due to the high relative weight of the (uniformly distributed) initial weights in early rounds, while the computer quantity in round 1 is not chosen by the learning theory, but set to 35, 40 or 45.
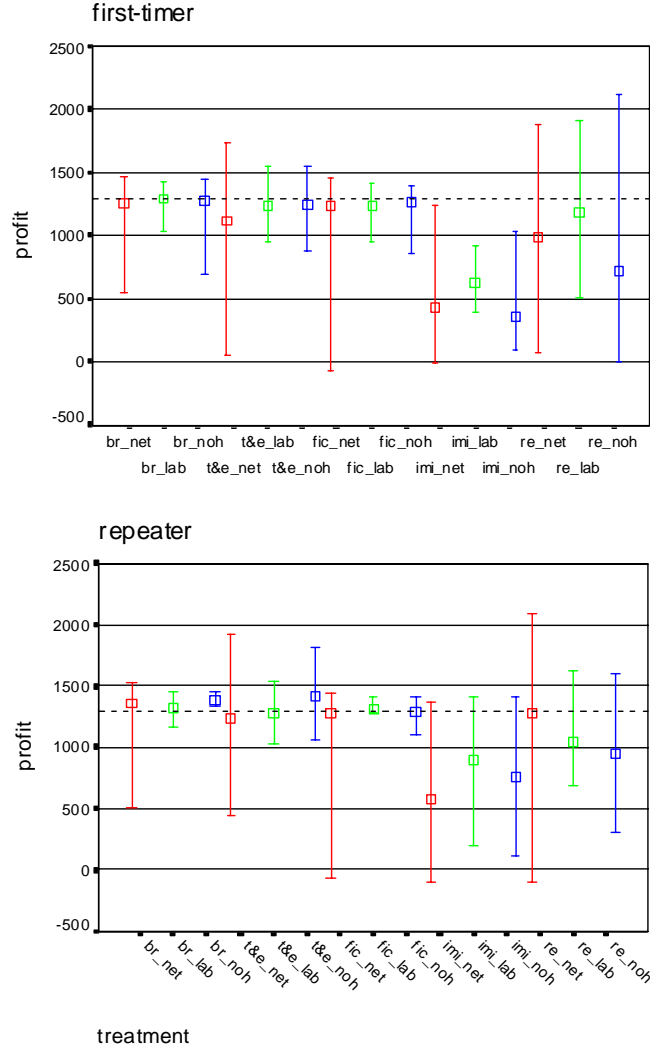
Figure 1: Range of human subjects' profits (first-timers and repeaters). The bars denote maximal, minimum, and, mean (the squares) profits for each treatment. The dashed line shows profit in the static Nash equilibrium. A treatment is a combination of computer opponent (br, t&e, fic, imi, re) and experimental setting (net, lab, noh).
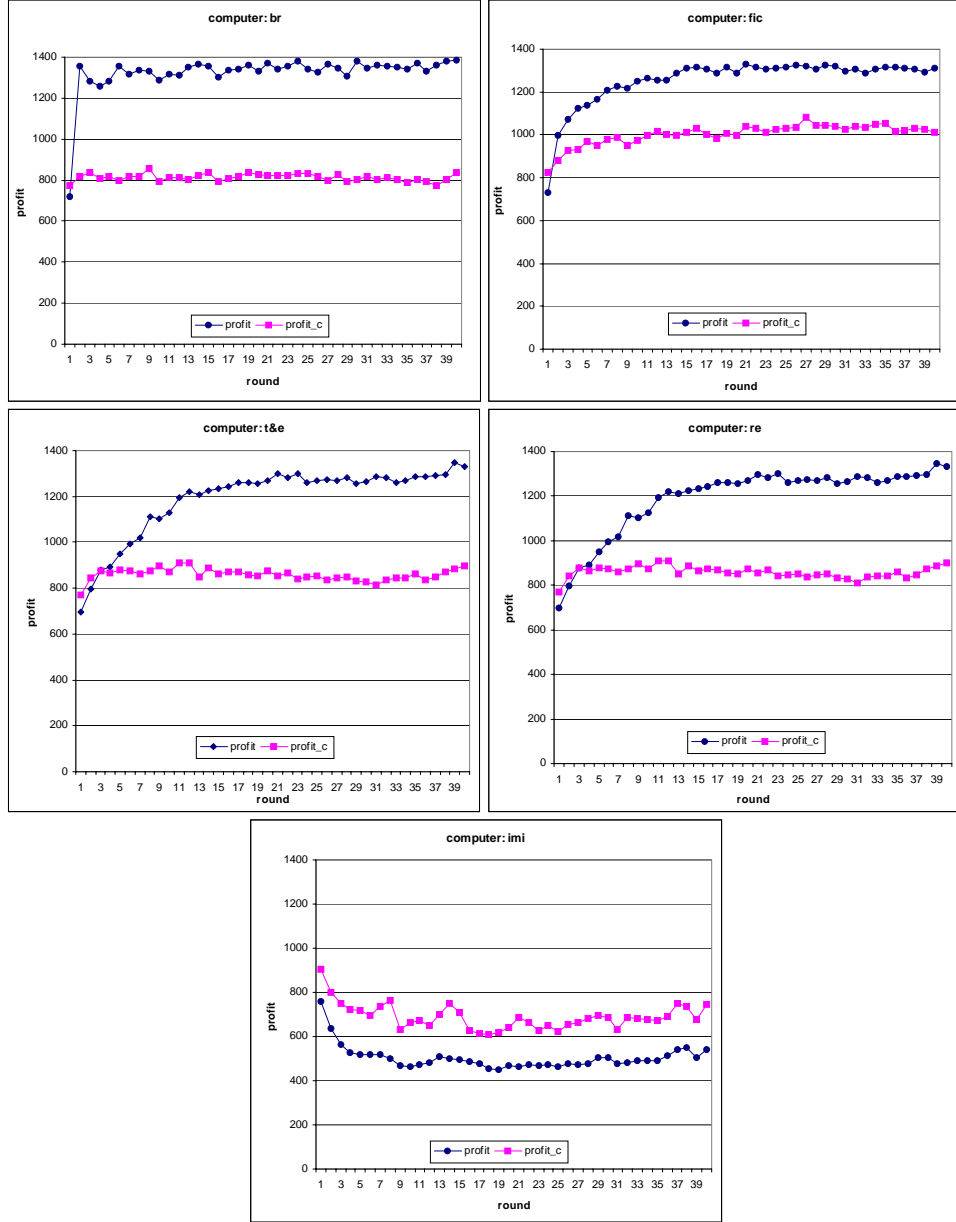
Figure 2: Time series of profits for subjects and computers for different computer types.

against a computer of type re, 27 who played against type t&e, and 21 who played against br. The top 10 players were almost exclusively playing against type re. This confirms the impression obtained from Figure 1. The highest profits can be achieved against type re but a lot of luck is needed for this due to the stochastic nature of reinforcement learning.

Table 5: Distribution of top subjects

against computer type...	among top 100	among top 10
br	21	–
t&e	27	1
re	52	9

Note: Pooled over all settings *net, lab, noh.*

5.2   Human tactics

In this section we want to describe human tactics, i.e., strategic teaching by subjects. We shall do so mainly by way of examples as it is difficult to reliably classify the behavior of all subjects. Least ambiguous is the classification of subjects playing against type br. Most of the top players in this treatment realized that profits are quite high when one stubbornly plays something close to the Stackelberg leader quantity.

Figure 3(a) shows quantities of the best subject playing against type br and the corresponding computer quantities. The best subject against br (ranked overall 57th) chose 55 in all 40 periods.[10] The computer quickly adjusted to a neighborhood of the Stackelberg follower quantity with the remaining movement due to the noise in the computer's decision rule.

Another interesting, though less frequent, pattern can be seen in Figure 3(b). The subject chose – with only slight variations – the following cycle of 4 quantities: 108, 70, 54, 42, 108, 70, ... Stunningly, this cycle produces an expected profit per round of 1520, which exceeds the Stackelberg leader profit.[11] By flooding the market with a quantity of 108, the subjects made

---

[10]Interestingly, none of our subjects chose the exact Stackelberg leader quantity of 54.

[11]The only reason the subjects in Figure 3(a) received an even higher payoff was luck

sure that the computer left the market in the next period. But instead of going for the monopoly profit, the subject accumulated intermediate profits over three periods. This, of course, raises the question, what is the optimal cycle? It turns out, that, in fact, the optimal cycle length is four and, after rounding to integers, the optimal cycle is 108, 68, 54, 41, which produces an expected profit of 1522. Thus, our subject was within 2 units of the solution for this non–trivial optimization problem.[12]

How did the very best subject play? Like all top players, he played against computer type re. Figure 4(a) reveals that the subject simply got lucky.[13] It was a first-time player in the no-history setting, i.e., a player with very little information about the game. The reinforcement algorithms locked in at very low quantities in the range of 10 and the subject roughly played a best response to that, which resulted in an average profit of 2117.

Finally, one could ask whether there were any successful attempts at collusion. Against computer types br, fic, and imi, collusion is theoretically impossible. Only for t&e there are theoretical results (Huck, Normann, and Oechssler, 2004a) which indicate that collusion could occur. However, as it turned out, the only successful example of collusion occurred against type re (see Figure 4(b)). Here the computer got locked in at about 27 and the subject consistently played 27. Of course, the subject could have improved his payoff by deviating to the Stackelberg leader quantity once the computer was locked in enough.

5.3 Can (myopic) learning theories describe subjects' behavior?

Are the myopic learning theories useful in describing subjects' behavior? In this section we analyze whether the same learning theories that were used to program the computers can be used to organize the behavior of their human

---

due to favorable noise of the computer algorithm.

[12] The subject played three times against br and left two comments. The first was "tried to trick him", the second "tricked him".

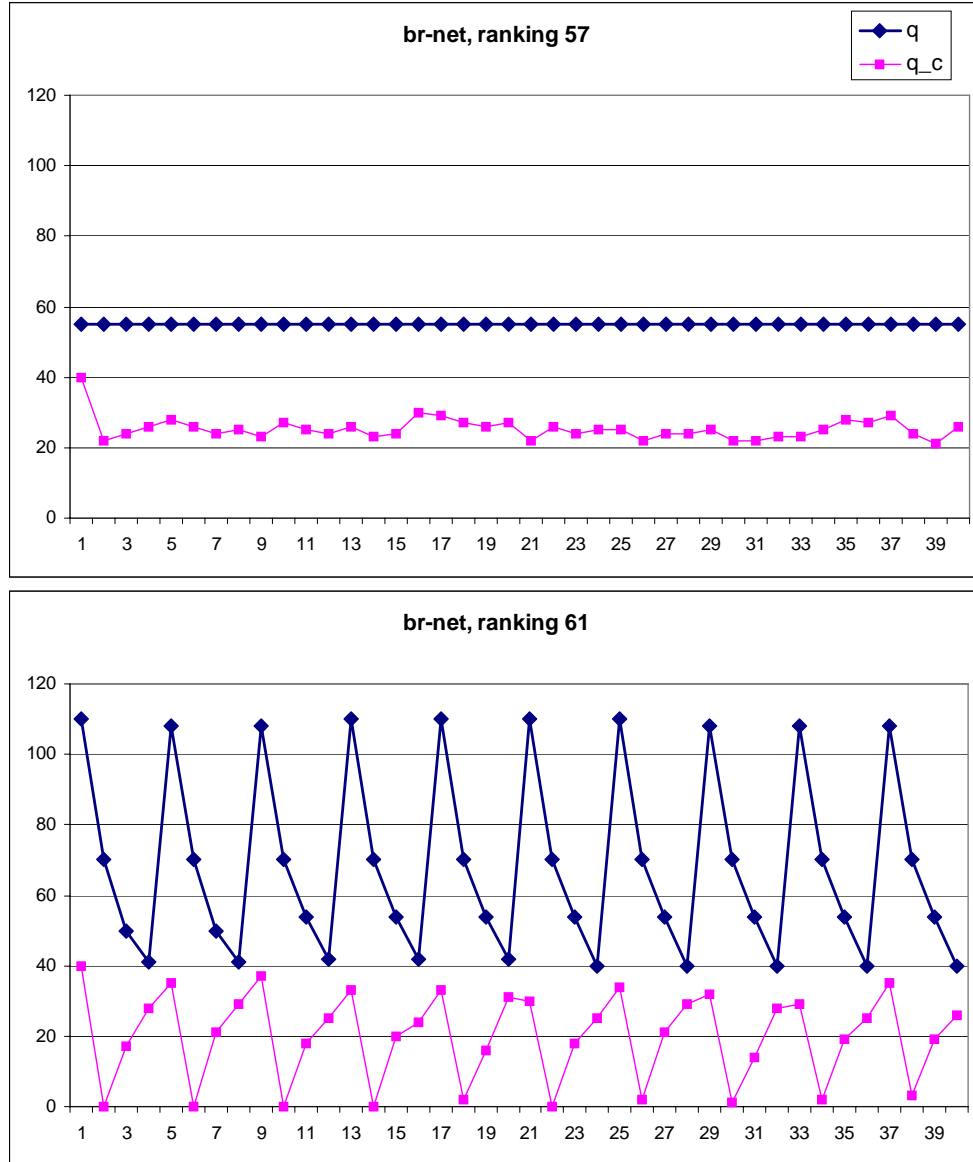[13] The description of his strategy was "π mal Daumen", which roughly translates to "rule of thumb".

Figure 3: (a) Quantities of subject ranked number 57 and of the br-computer opponent (top panel); (b) Quantities of subject ranked number 61 and of the br-computer opponent (lower panel)
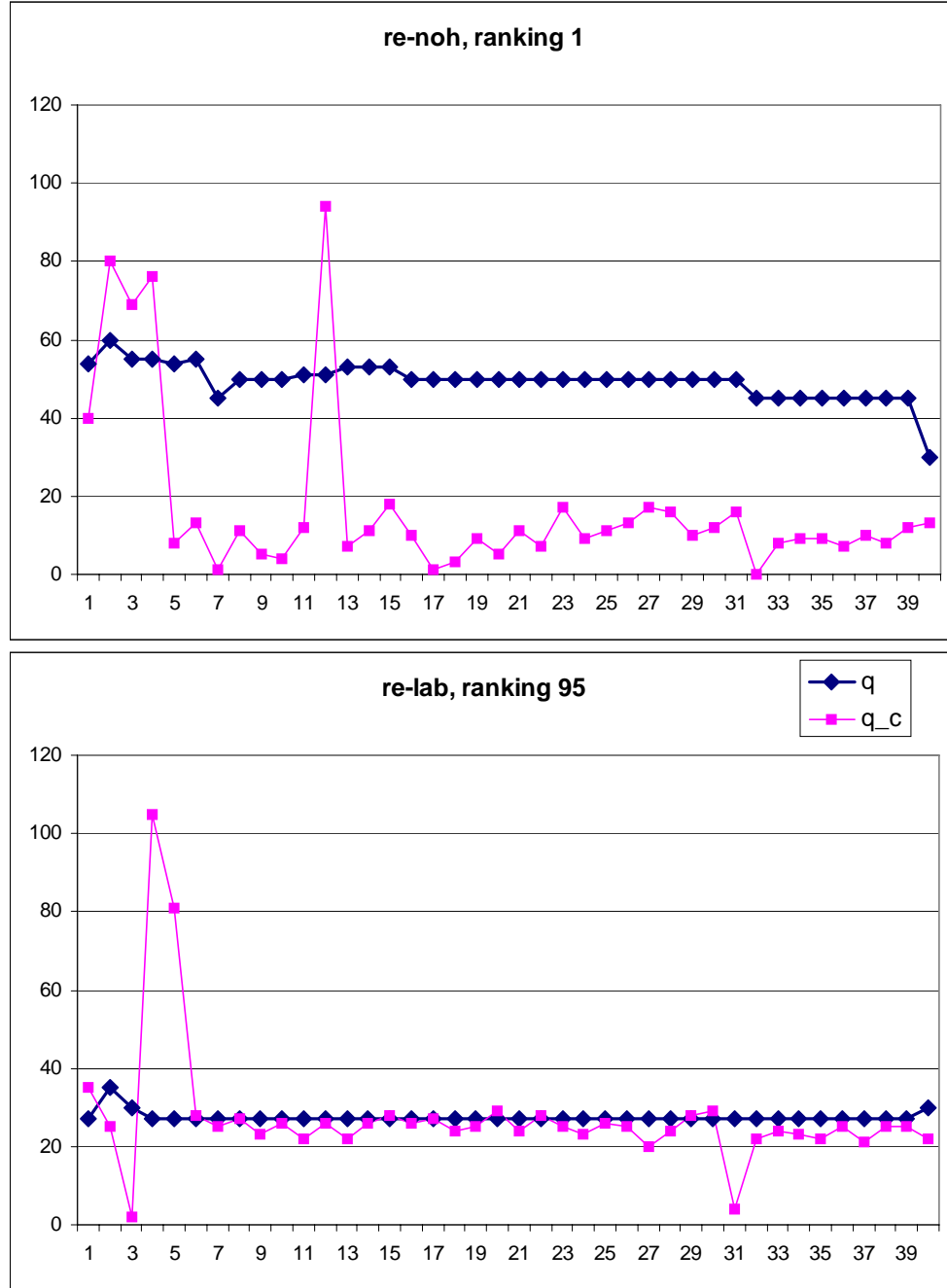
21

Figure 4: (a) Quantities of top-ranked subjects and his re-computer opponent (top panel); (b) Quantities of a pair that managed to achive collusion (lower panel).

opponents. We shall do so by calculating for each round (except round 1) the quantity q_i^*, which is predicted by the respective theory (without noise), and comparing it to the actually chosen quantity in that round, q_i^t.[14] The mean squared deviation (MSD), $(q_i^* - q_i^t)^2$, is then calculated for each theory by averaging over all periods $t = 2, ..., 40$, all subjects, and all treatments. We also calculate MSDs for the predictions of constant play of the Stackelberg leader quantity, for constant play of the Cournot Nash equilibrium quantity, for constant play of the collusive quantity, and for simply repeating the quantity decision from the previous round ("same"). Finally, as a benchmark we calculate the MSD that would result from random choice generated by an i.i.d. uniform distribution on [0,109] ("random"). Figures 5 and 6 show the resulting average MSD for the settings *net* and *lab* separately. Both figures demonstrate that all predictions perform substantially better than random choice. Reinforcement has the lowest MSD, followed by trial&error, same, and imitation. Not surprisingly, collusion is very far off the mark. A similar picture emerges for both experimental settings except that MSDs in the lab are generally much lower than in the internet setting. It seems that subjects in the lab are better described by our theories.

A slightly different ranking of learning theories is obtained when we consider the theory that best describes a subject's play (measured by minimum MSD for all decisions of a given subject in periods 2 through 40). Figure 7 lists the number of subjects' plays that are best described by the various theories. Here imitation is most frequently the best fitting theory. Overall, we see that the myopic learning theories do have some descriptive power. Yet, given the observed tendency of subjects towards strategic teaching, we should not be surprised that the fit is all but perfect.

5.4   A comparison with human vs. human data

It should be interesting to compare the behavior of our subjects to that of subjects in a "normal" experiment where subjects play against other humans subjects instead of computers. For this purpose we look at the duopoly

---

[14]In the case of reinforcement learning we take q_i^* to be the expected quantity given the distribution of propensities.
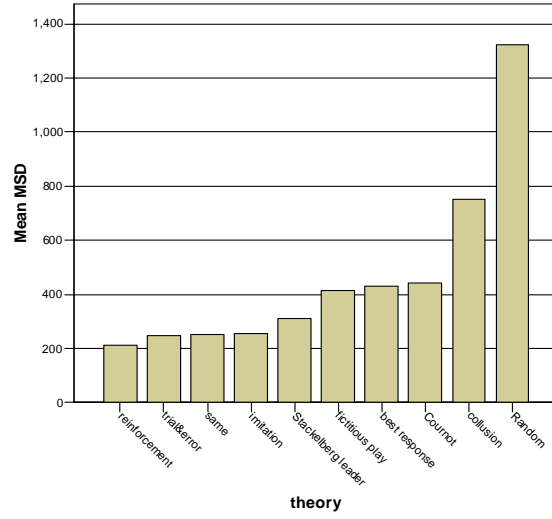
Figure 5: Average MSD for various theoretical predictions, setting net
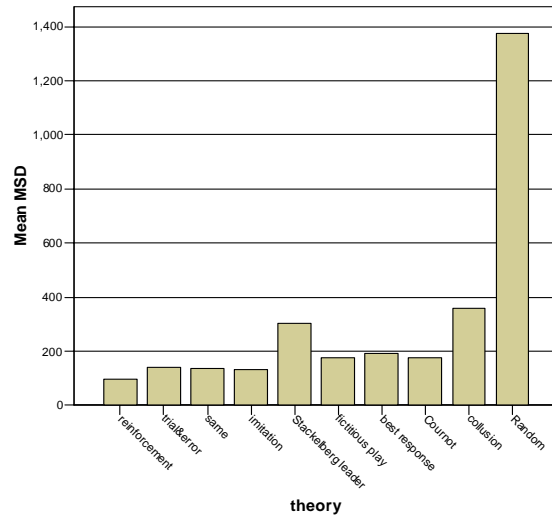Note: Average is taken over all periods, all subjects, and all treatments.

Figure 6: Average MSD for various theoretical predictions, setting lab
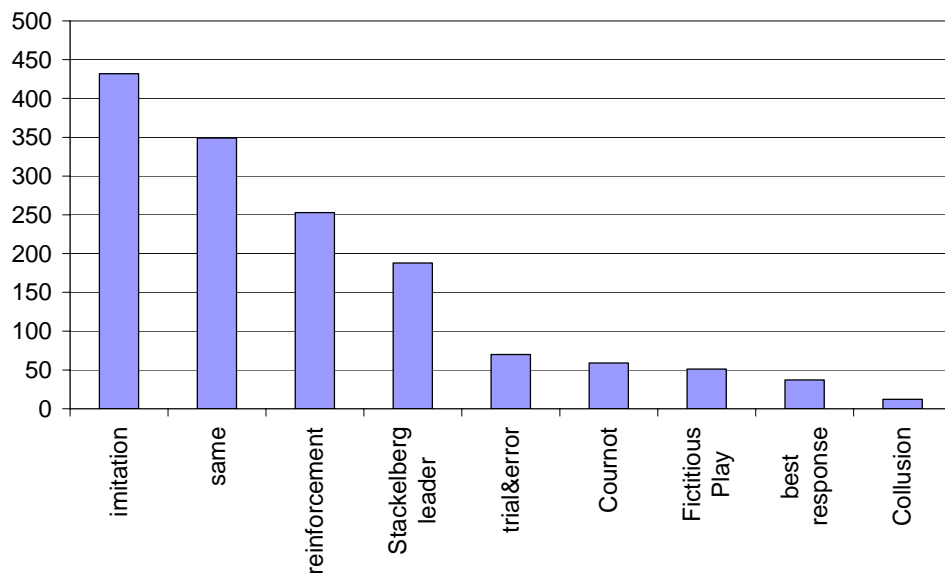Note: Average is taken over all periods, all subjects, and all treatments.

Figure 7: Number of plays best described by the various theoretical predictions, all settings

Note: A theory is said to best describe a subject's play if it minimizes MSD over periods 2-40 of that play.

treatment of Huck, Normann, and Oechssler (2004b), which has a fairly similar design as the current experiment.[15] A striking difference in results appears when we compare the average quantities of human subjects. While in Huck, Normann, and Oechssler (2004b) the average quantity of (human) subjects is about 9% below the Nash equilibrium quantity, it is more than 33% above the Nash quantity for human subjects in the current experiment (more than 17% above Nash in *lab*). That is, when subjects know that their opponents are also human subjects, they behave slightly collusive. When they know that they play against computers, they play substantially more aggressively.

---

[15] The main differences in the design are that Huck, Normann, and Oechssler (2004b) use a demand function with $p = 100 - q_i - q_{-i}$ and a finer grid of strategies. Furthermore, their experiment lasted for only 25 periods. All other design features were essentially the same.

As in the previous subsection we can calculate the MSD for how well the studied learning theories describe humans' behavior. Figures 8 and 9 show the average MSD for each of the learning theories best response, fictitious play, imitation, reinforcement learning, and trial & error, for our human vs. computer experiment and Huck et al.'s (2004b) human vs. human experiment, respectively. Note that the levels of the MSD for the two experiments are not perfectly comparable since the demand functions differ slightly. Nevertheless, it is striking how much lower the average MSD are for the human vs. human experiment. In any case, the ranking of the different learning theories in terms of MSD is informative, and this ranking is almost exactly reversed: those theories that describe the human behavior best in our experiment, namely best response and fictitious play, turn out to be those that describe human behavior worst in the context of a human vs. human situation.

When we look at the average MSD for all 5 learning theories, we see that descriptive power of those theories becomes better over time. Figures 10 and 11 show the development of average MSD for all theories, separately for our human vs. computer experiment and Huck et al's (2004b) human vs. human experiment. While there is improvement for both experiments, the improvement in the human vs. human case is much stronger.

What could account for this? Both, best response and fictitious play work well when describing play near a Cournot equilibrium. Looking at Table 4, we see that subjects are more likely to play quantities close to the Cournot equilibrium when playing against other humans, and consequently are better described by fictitious play and best response. But why does this not apply when playing against a computer? It seems that strategic teaching is more pronounced when playing against a computer. Strategic teaching usually consists of playing higher quantities to induce the computer to react with lower quantities in future rounds. Since such forward-looking behavior is not predicted by any of the five (adaptive) learning theories, average MSDs remain relatively high in the human vs. computer experiment.

Reasons for the subjects to use less strategic teaching against other humans could include fairness considerations (the Cournot outcome is "fairer"
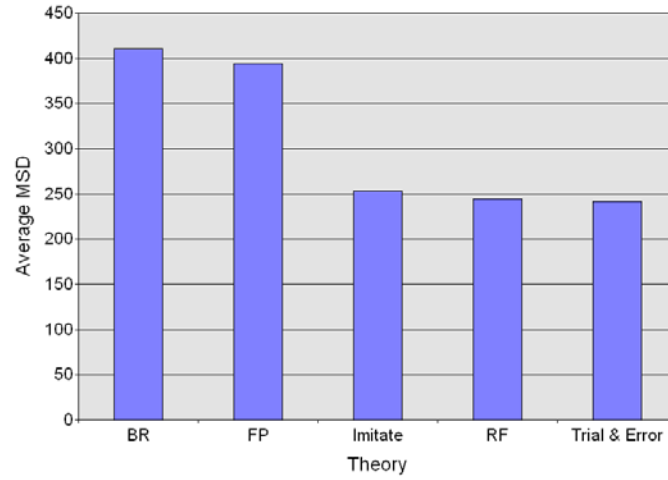
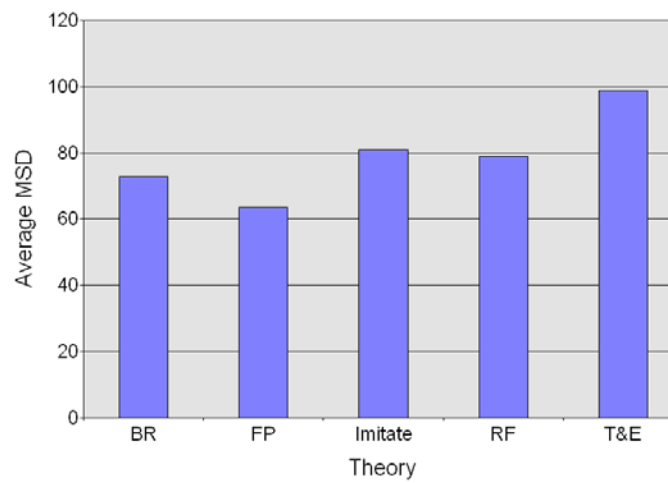Figure 8: Average MSD of different learning theories in human vs. computers experiment

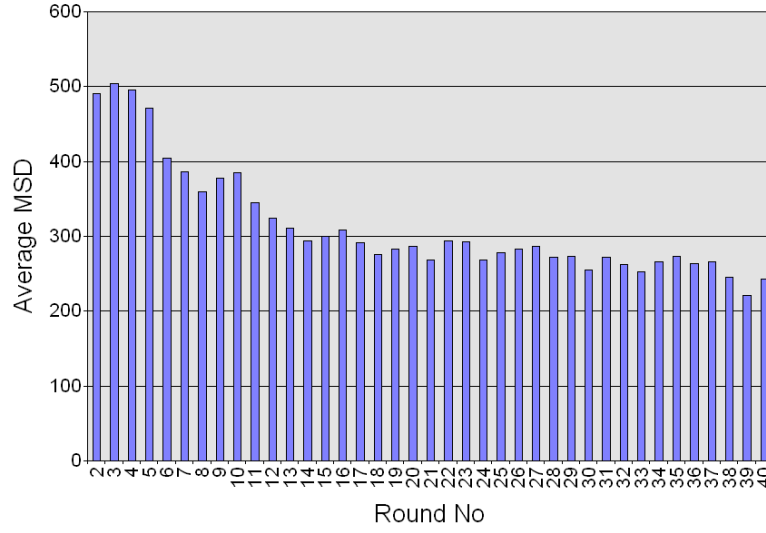Figure 9: Average MSD of learning theories in human vs. human experiment (Data from Huck et al. 2004b)

Figure 10: Average MSD of all 5 learning theories over time, human vs. computer experiment
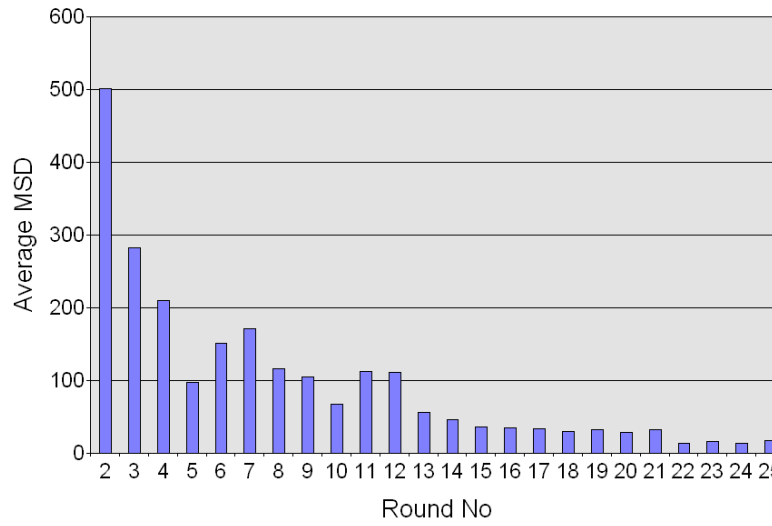

Figure 11: Average MSD of all 5 learning theories over time, human vs. human experiment

28

than the Stackelberg outcome) and the anticipation of negative reciprocal reactions. Alternatively, subjects may believe that real subjects are harder to fool (or more stubborn) than a simple computer program and are therefore less susceptible to strategic teaching. Of course, there is no good reason for supposing that computers could not be programmed to mimic "emotional" reactions of humans like reciprocity, revenge, or, indeed, rage. But probably our dominant perception of computers is one of rationally acting machines without emotions.

5.5   Learning theories and economic value

We define the economic value of a given learning theory as the improvement in a subject's profit generated by substituting the learning theory's recommendation for the actual choice of the subject (compare Camerer and Ho, 2001), where the learning theory's choice is based on the real history of play up to that round. Note that this is of course a very myopic point of view: only improvements in payoffs for the current period are counted whereas possible long–term gains are ignored.

Figure 12 shows the average economic value that the five different learning theories would have generated for our subjects, separately for the three experimental settings *lab*, *net*, and *noh*, and the five learning theories. While the economic value of "imitate the best" is rather low and that of "trial & error" even negative, there are substantial potential gains from switching to best response, fictitious play, or reinforcement learning, considering that the average profit per round was about 1112. Figure 12 shows that the ranking of the learning theories in terms of economic value are very similar across experimental settings, but the levels are lower in *lab*. Just as our subjects in *lab* are better described by the learning theories, the additional value of having those learning theories' advice is reduced.

As pointed out above, it should not come as a surprise that the economic values are so high despite the fact that subjects actually achieved much higher profits than computers. Since economic value does not capture the long-term effects of a strategy, it does not capture strategic teaching. As we
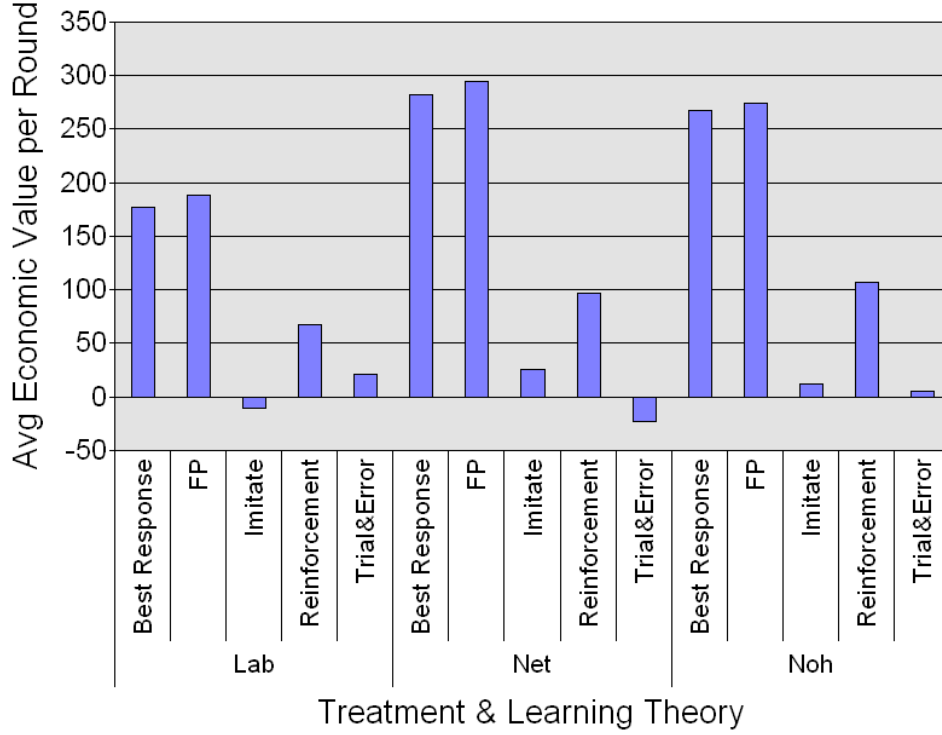
Figure 12: Economic value of different learning theories, separately for different experimental settings lab, net, and noh.

saw in Section 5.2, quite a number of our subjects were successfully trying to exploit the learning theories' algorithms by deliberately foregoing profits in the current round to induce the computer opponent to play in a way that enables the human to gain larger profits in future rounds. Thus, high economic value may just be a sign that a subject is deliberately deviating from the myopic optimum to maximize long-term profits.

5.6 Experimenting on the internet - does it make a difference?

Looking at Table 4 it is apparent that subjects' average quantities on the net seem to be substantially higher than in the lab. In fact, when we aggregate the mean quantities shown in Table 4 over computer types, we get average

quantities of 48.68 in *net* and 42.30 in *lab*. This difference is significant at all conventional significance levels for *t*-tests or Mann-Whitney U tests. What does account for this difference?

If this difference were driven by the different environment (internet versus laboratory), this would be problematic for the future use of internet experiments. Note, however, that our *net* and *lab* settings differ also by other aspects, in particular the incentive scheme and possibly the subject pool. In *lab*, we paid subjects according to their performance. In *net*, subjects were solely motivated by their ranking on the highscore table.[16] Note that this difference is *not* about relative (*net*) versus absolute (*lab*) payoff maximization. A subjects needs to maximize his absolute payoff in order to achieve a large highscore.

To sort those things out, we have conducted experiments with two additional settings. The two new settings are designed to bridge the gap between the *lab* and *net* settings. The setting "*lab*-f" is just like the *lab* setting except that subjects received a *fixed* payment of 10 Euros as soon as they entered the *lab*.[17] Setting "*lab*-np" is like *lab* except that subjects received *no* payment at all. Thus, in both new settings, a good placement on the highscore table was the only motivation for subjects. The only difference between *lab*-np and *net* was the environment, that is, the laboratory versus subjects' homes or offices. To summarize, the new and old settings can be ordered as follows.

$$lab \quad \overset{\text{fixed pay}}{\longrightarrow} \quad lab\text{-f} \quad \overset{\text{no pay}}{\longrightarrow} \quad lab\text{-np} \quad \overset{\text{lab vs. home}}{\longrightarrow} \quad net \qquad (5)$$

The experiments for setting *lab*-f were conducted in October 2004 in the Bonn Laboratory of Experimental Economics. There were 50 subjects who each played twice against the same computer type, just like in setting *lab*. Subjects for setting *lab*-np were volunteers who took part in an introduction for freshmen during which they visited the laboratory. There were 55

---

[16] For some subjects getting the top-spot on a highscore table presents substantial incentives. For at least one subject the incentive was so great that he or she invested sufficient time to hack our system, and tried to manipulate the highscore table.

[17] In principle, subjects could have left the lab after receiving the 10 Euros but no one did.

volunteers of which 5 played a second time.

Each of the arrows in (5) could account for the difference in quantities between *lab* and *net*. Table 6 shows mean quantities for the different setting for first-time players and all subjects, separately.

Table 6: Mean quantities

setting	first-timers' mean quantities	all subjects' mean quantities
lab	43.14	42.30
lab-f	48.21	47.65
lab-np	48.85	48.52
net	48.69	48.68

Note: Average quantities over all 40 rounds.

Table 6 shows clearly that there is a significant difference (p–values < 0.001 for t-tests and Mann-Whitney U tests) only for the first of those arrows, i.e. between *lab* and *lab*-f. There are no significant differences at any conventional level between *lab*-f, *lab*-np, and *net*. We conclude that the difference between *lab* and *net* is primarily driven by the lack of monetary incentives in *net* and not by the environment of the decision maker.

6   Conclusion

In this experiment we let subjects play against computers which were programmed to follow one of a set of popular learning theories. The aim was to find out whether subjects were able to exploit those learning algorithms. The bulk of the (boundedly rational) learning theories that have been studied in the literature (see Fudenberg and Levine, 1998, for a good overview) are *myopic* in nature. Probably the most fundamental insight from our experiment is that we need to advance to theories that incorporate at least a limited amount of foresight. Many of our subjects were quite able to exploit the simple myopic learning algorithms. *Strategic teaching* is an important phenomenon that needs to be accounted for in the future development of theory. Yet, a word of caution is in order. A comparison with human vs. human data reveals that myopic learning theories are much better able to

explain behavior than in our human vs. computer experiment. Why this is so, remains an interesting question for future work.

Our experiment also provides some methodological lessons with respect to internet experiments. Although we found significant differences between our internet and our laboratory setting, we could fully account for those differences through the different incentive schemes. Internet experiments are fine, as long as subjects have proper monetary incentives.

Appendix

A Instructions

A.1 Introduction Page Internet

Welcome to our experiment!

Please take your time to read this short introduction. The experiment lasts for 40 rounds. At the end, there is a high score showing the rankings of all participants. You represent a firm which produces and sells a certain product. There is one other firm that produces and sells the same product. You must decide how much to produce in each round. The capacity of your factory allows you to produce between 0 and 110 units each round. Production costs are 1 per unit. The price you obtain for each sold unit may vary between 0 and 109 and is determined as follows. The higher the combined output of you and the other firm, the lower the price. To be precise, the price falls by 1 for each additional unit supplied. The profit you make per unit equals the price minus production cost of 1. Note that you make a loss if the price is 0. Your profit in a given round equals the profit per unit times your output, i.e. profit = (price 1) * Your output. Please look for an example here. At the beginning of each round, all prior decisions and profits are shown. The other firm is always played by a computer program. The computer uses a fixed algorithm to calculate its output which may depend on a number of things but it cannot observe your output from the current round before making its decision. Your profits from all 40 rounds will be added up to calculate your high score. There is an overall high score and a separate one for each type of computer. Please do not use the browser buttons (back, forward) during the game, and do not click twice on the go button, it may take a short while.

Choose new quantity

Please choose an integer (whole number) between 0 and 110.

A.2 Introduction Page *lab*

Welcome to our experiment!

Please take your time to read this short introduction. The experiment lasts for 40 rounds. Money in the experiment is denominated in Taler (T). At the end,

exchange your earnings into Euro at a rate of 9.000 Taler = 1 Euro. You represent a firm which produces and sells a certain product. There is one other firm that produces and sells the same product. You must decide how much to produce in each round. The capacity of your factory allows you to produce between 0 and 110 units each round. Production cost are 1T per unit. The price you obtain for each sold unit may vary between 0 T and 109 T and is determined as follows. The higher the combined output of you and the other firm, the lower the price. To be precise, the price falls by 1T for each additional unit supplied. The profit you make per unit equals the price minus production cost of 1T. Note that you make a loss if the price is 0. Your profit in a given round equals the profit per unit times your output, i.e. profit = (price 1) * Your output. Please look for an example here. At the beginning of each round, all prior decisions and profits are shown. The other firm is always played by a computer program. The computer uses a fixed algorithm to calculate its output which may depend on a number of things but it cannot observe your output from the current round before making its decision. Your profits from all 40 rounds will be added up to calculate your total earnings. Please do not use the browser buttons (back, forward) during the game, and do not click twice on the go button, it may take a short while.

Choose new quantity

Please choose an integer (whole number) between 0 and 110.

A.3 Example Page

The Formula

The profit in each round is calculated according to the following formula:

Profit = (Price 1) * Your Output

The price, in turn, is calculated as follows.

Price = 109 Combined Output

That is, if either you or the computer raises the output by 1, the price falls by 1 for both of you. (but note that the price cannot become negative). And the combined output is simply:

Combined Output = Your Output + Computers Output

Example:

Lets say your output is 20, and the computers output is 40. Hence, combined output is 60 and the price would be 49 (= 109 - 60). Your profit would be (49 1)*20 = 960. The computers profit would be (49 - 1)*40 = 1920. Now assume you raise your output to 30, while the computer stays at 40. The new price would be 39 (= 109-40-30). Your profit would be (39 - 1)*30 = 1140. The computers profit would be (39 - 1)*40 = 1520.

To continue, please close this window.

B   Screenshots

The Game

Welcome to our experiment!

Please take your time to read this short introduction.

The experiment lasts for 40 rounds. At the end, there is a highscore showing the rankings of all participants. You represent a firm which produces and sells a certain product. There is **one** other firm that produces and sells the same product. You must decide how much to produce in each round. The capacity of your factory allows you to produce between 0 and 110 units each round. Production cost are 1 per unit.

The price you obtain for each sold unit may vary between 0 and 109 and is determined as follows. The higher the combined output of you and the other firm, the lower the price. To be precise, the price falls by 1 for each additional unit supplied.

The profit you make per unit equals the price minus production cost of 1. Note that you make a loss if the price is 0. Your profit in a given round equals the profit per unit times your output, i.e. profit = (price − 1) * Your output. Please look for an example here. At the beginning of each round, all prior decisions and profits are shown.

The other firm is always played by a **computer program**. The computer uses a fixed algorithm to calculate its output which may depend on a number of things but it cannot observe your output from the current round before making its decision.

Your profits from all 40 rounds will be added up to calculate your high score. There is an overall high score and a separate one for each type of computer.

Please **do not use the browser buttons** (back, forward) during the game, and do not click twice on the go button, it may take a short while.

β

You are playing beta

How to play

Enter a number between 0 and 110 in the field and click go.

More Informations

- Introduction
- The formula

Choose new quantity

Please choose an integer (whole number) between 0 and 110.

[   ] Go!

History

Round Your quantity Your profit Computer quantity Computer profit

no history yet

Choose new quantitity

Please choose an integer (whole number) between 0 and 110.

[   ] Go!

β

You are playing beta

History

Round	Your quantity	Your profit	Computer quantity	Computer profit
14	34	1360	34	1360
13	45	2205	14	686
12	75	-75	39	-39
11	27	1242	35	1610
10	38	1406	33	1221
9	48	1440	30	900
8	47	1551	28	924
7	46	1380	32	960
6	45	1350	33	990
5	45	1530	29	986
4	44	1540	29	1015
3	44	1408	32	1024
2	44	1452	31	1023
1	44	1276	35	1015

How to play

Enter a number between 0 and 110 in the field and click go.

More Informations

- Introduction
- The formula

Round number 1 of 40

Round	Your quantity	Your profit	Computer quantity	Computer profit
1	44	1276	35	1015

go on...

38

References

[1] Alós–Ferrer, C. (2004). Cournot vs. Walras in dynamic oligopolies with memory, International Journal of Industrial Organization, 22, 193-217.

[2] Apesteguia, J., Huck, S. and Oechssler, J. (2003). Imitation - Theory and experimental evidence, University of Bonn.

[3] Brown, G.W. (1951). Iterative solutions of games by fictitious play, in: Koopmans, T.C. (ed.), Activity analysis of production and allocation, John Wiley.

[4] Camerer, C., and Ho, T.H. (2001). Strategic learning and teaching in games, in: S. Hoch and H. Kunreuther (eds.) Wharton on decision making, New York: Wiley.

[5] Camerer, C., Ho, T.H. and Chong, J.K. (2002). Sophisticated experience-weighted attraction Learning and strategic teaching in repeated games, Journal of Economic Theory, 104, 137-188.

[6] Coricelli, G. (2005). Strategic interaction in iterated zero-sum games, Homo Oeconomicus, forthcoming.

[7] Cournot, A. (1838). Researches into the mathematical principles of the theory of wealth, transl. by N. T. Bacon, MacMillan Company, New York, 1927.

[8] Drehmann, M., Oechssler, J. and Roider, A. (2005). Herding and contrarian behavior in financial markets, American Economic Review, forthcoming.

[9] Ellison, G. (1997). Learning from personal experience: One rational guy and the justification of myopia, Games and Economic Behavior, 19, 180-210.

[10] Erev, I. and Roth, A. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria, American Economic Review, 88, 848-881.

[11] Fox, J. (1972). The learning of strategies in a simple, two-person zero-sum game without saddlepoint, Behavioral Science, 17, 300-308.

[12] Fudenberg, D., and Levine, D. (1998). The theory of learning in games, Cambridge: MIT Press.

[13] Houser, D. and Kurzban, R. (2002). Revisiting kindness and confusion in public goods experiments, American Economic Review, 94, 1062-1069.

[14] Huck, S., Normann, H.T., and Oechssler, J. (1999). Learning in Cournot oligopoly: An experiment, Economic Journal, 109, C80-C95.

[15] Huck, S., Normann, H.T., and Oechssler, J. (2004a). Through trial & error to collusion, International Economic Review, 45, 205-224.

[16] Huck, S., Normann, H.T., and Oechssler, J. (2004b). Two are few and four are many: Number effects in experimental oligopoly, Journal of Economic Behavior and Organization, 53, 435-446.

[17] Ianni, A. (2002). Reinforcement learning and the power law of practice: Some analytical results, University of Southampton.

[18] Kirchkamp, O. and Nagel, R. (2005). Naive learning and cooperation in network experiments, mimeo, Universitat Pompeu Fabra.

[19] Laslier, J.-F., Topol, R. and Walliser, B. (2001). A behavioral learning process in games, Games and Economic Behavior, 37, 340-366.

[20] Lieberman, B. (1962). Experimental studies of conflict in some two-person and three-person games, in: Criswell, J. H., Solomon, H. and Suppes, P. (eds.), Mathematical methods in small group processes, Stanford University Press, 203-220.

[21] Matros, A. (2004). Simple Rules and Evolutionary Selection, University of Pittsburgh.

[22] McCabe, K., Houser, D., Ryan, L., Smith, V. and Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange, Proceedings of the National Academy of Sciences, 98, 11832-11835.

[23] Messick, D.M. (1967). Interdependent decision strategies in zero-sum games: A computer controlled study, Behavioral Science, 12, 33-48.

[24] Monderer, D. and Shapley, L. (1996). Potential games, Games and Economic Behavior, 14, 124-143.

[25] Offerman, T., Potters, J., and Sonnemans, J. (2002). Imitation and belief learning in an oligopoly experiment, Review of Economic Studies, 69, 973-997.

[26] Robinson, J. (1951). An iterative method of solving games, Annals of Mathematics, 54, 296-301.

[27] Roth, A. and Erev, I. (1995). Learning in extensive form games: Experimental data and simple dynamic models in the intermediate term, Games and Economic Behavior 8, 164-212

[28] Roth, A. and Schoumaker, F. (1983). Expectations and reputations in bargaining: An experimental study, American Economic Review, 73, 362-372.

[29] Sarin, R. and Vahid, F. (2004). Strategic similarity and coordination, Economic Journal, 114, 506-527.

[30] Schipper, B.C. (2004), Imitators and optimizers in Cournot oligopoly, University of Bonn.

[31] Shachat, J. and Swarthout, J. T. (2002). Learning about learning in games through experimental control of strategic independence, University of Arizona.

[32] Thorndike, E.L. (1898). Animal intelligence: An experimental study of associative processes of animals, Psychological Monographs, 2 (8).

[33] Vega–Redondo, F. (1997). The evolution of Walrasian behavior, Econometrica, 65, 375-384.

[34] Walker, J., Smith, V.L. and Cox, J.C. (1987). Bidding behavior in first price sealed bid auctions, Economics Letters, 23, 239-244.